



Summarization of Interest Group Outcomes

Beth Plale and *Peter
Wittenberg*

(presentation at rd-alliance.org)

Assessment criteria

- RDA Working Group profile:
 - *Short-term: 12-18 months*
 - *Focused efforts with specific actions adopted by specific communities*
 - *International participation*
 - *Open, voluntary, consensus-driven*
 - *Complementary to effective efforts elsewhere*
- *Potential outcomes/deliverables:*
 - *Adopted standards*
 - *Deployed infrastructure*
 - *Adopted policy*
 - *Implemented best practice, etc*
- *IMPACT:*
 - *Are deliverables and WG Action Plan achievable and do they produce tangible results?*
 - *Are there measurable outcomes?*
 - *Will product(s) of Working Group foster data sharing and/or exchange?*

Discussion Points in the Data Foundation Session

In this note it is briefly reported which kind of topics were raised in the 2 presentation rounds and which topics converged into actions already. Please, comment when something is missing.

- Terminology Harmonization
- Metadata – direction minimal set
- Policies
- Long term Preservation
- XLBS/XLBM

Other topics which were mentioned and which were not turned to concrete suggestions: data management, provenance management, relations between stakeholders, how to get scientists involved, training, semantic interoperability, generic bridge, vocabularies, access to long-tail data, AAI, data science, ethics of data sharing, protocols to enable Data as a Service, endpoints for RDA, extensive/powerful cross-disciplinary data model

Notes per Topic

Here the points are listed in telestyle which were mentioned during the discussion of the topics which were turned into suggestions. They are certainly not comprehensive. Ordering is roughly by number of supporters in the session.

Terminology Harmonization

basic concepts of data world and organizations, basic IT principles, necessity to overcome understanding difficulties in cross-disciplinary work, definitions in open registries, reference to best practices, reference to implementations
outcome should be a reference guide

Metadata

minimal metadata for scientific data, overview of metadata schemas and concepts, understand needs of researchers, discovery is relevant, basics to be adhered to (enter PIDs to data, register schemas, register concepts), support for life cycle of metadata, layers/function of metadata, validation of md, lightweight methods to add metadata, provenance aspects
outcome should be a minimal set and best practice guide

Notes per Topic

Policies

examples for policy sets from different communities, “layers” of policies, provenance related with policy rule execution, classification of policies, use of policies for quality assessment, result of quality assessment in metadata, quality of data as an overarching issue

outcome should be best practice guide with good examples

Long-term Preservation

– was not so much time for discussion

relevance of policies for proper ltp, get data on the “data highway”, best practice use cases

outcome should be best practice guide with good examples

XLDB/XLDM (Editor: Yannis) – was not so much time for discussion

management of very large data sets, data management from large sensor arrays, organizing large data sets in several communities

outcome should be use cases, best practices, benchmarks, standardization, special policies

Discussion points in registries session

- In this note reports about our understanding of the issues mentioned. Please, comment when something is missing.
- Topics that were turned into short descriptions:
 - Data Center Registries
 - Data Type Registries
 - PID Information Type
 - Outsourcing and Process Automation
- Other topics which were mentioned: other types of registries, registry interchange, data as a service

Notes per Topic

Data Centre Registries. Maintained interrogate-able list of current data centers and what they offer. Group saw value in data centre registry; especially one that could address “best deposit” problem.

- Needs more time to refine problem statement
- Would DataOne use this? (see Rebecca’s panel statement.)

Type Registries. All data is typed, explicitly or implicitly and processing requires knowledge of types. Be able to look up type for data object; find services that operate on that type.

- How demonstrate demand? Interest by which DCs?
- How keep registry current?
- Outcome of WG?

Notes per Topic

PID Information Type Harmonization. What information should be associated with PIDs? Core set or complete set ? Getting existing systems to adopt this; getting DCs to use them.

- Right people in discussion? (e.g., data center folks)
- Barriers to adoption?
- Proposed deliverables?

Outsourcing and Process Automation. Flickr, Dropbox show how simple outsourced services can liberate data by automating tasks relating to data storage, sharing management, etc. Software-as-a-service has power to reduce friction of data sharing, and reduce data center costs (through outsourcing).

- what are the adoptable ideas from this effort?

Discussion Points in the “Various” Session

In this note reports about our understanding of the issues mentioned.
Please, comment when something is missing.

Topics that were turned into short descriptions:

- Certification of Repositories
- Linking Data and Publications
- Education Workforce Issues
- Legal Interoperability including IPR & confidentiality

Other topics which were mentioned: cataloging existing efforts, preservation, communications, policy, outreach, sustainability, authentication/access, interop in marine and medical data

Notes per Topic

Certification of Repositories

- need more awareness for trust building on repositories worldwide
- implementation and funding of assessment practices worldwide
- there are BP suggestions
- outcome should be worldwide standard/best practice for certification?
- accelerant role, effect on sharing data, lack of trust hampers sharing

Linking Data and Publications

- more some questions of how to tackle this
- obviously a very relevant point
- outcome should be catalogue of BP, acceptance wrt to career building

Education Workforce Issues

- questions/overview about role of stakeholders, incentives, curricula
- obviously a very relevant point
- outcome should be ?
- competencies of actors, consistencies in this

Notes per Topic

Legal Interoperability including IPR & confidentiality

- better understanding of public law and licensing for use/reuse
- better mechanisms for integrating “legal policy compatible” data
- outcome should be extension/comment on Creative Commons Model and BP guideline to improve data sharing